

REMARKS

Reconsideration and allowance in view of the foregoing amendment and the following remarks are respectfully requested. Claims 1, 13 and 21 are amended without prejudice or disclaimer. Dependent claims 2-12, 14-20 and 22-32 are amended for consistency with respective amended parent claims.

Applicants thank the Examiner for the detailed response in the current Office Action as well as making the present Office Action non-final in order to provide Applicants with the opportunity to further consider the remarks and provide a response. Applicants have amended the claims and shall provide a detailed explanation of why the prior art fails to teach the value associated with the probability of the user being in a particular acoustic environment.

Rejection of Claims 1-4, 8-9, 13, 15, 18-19, 21-23 and 27-29 Under 35 U.S.C. §103(a)

The Office Action rejects claims 1-4, 8-9, 13, 15, 18-19, 21-23 and 27-29 under 35 U.S.C. §103(a) as being unpatentable over Komori et al. (U.S. Patent No. 7,050,974) ("Komori et al.") in view of Besling et al. (U.S. Patent No. 6,363,348) ("Besling et al."). Applicants have amended claims 1, 13 and 21 and respectfully submit that these claims are patentable and in condition for allowance. Applicants shall focus on the limitation of the memory that stores a user profile having data related to user vocal information and a value associated with a probability of the user being in a particular acoustic environment.

Applicants do not concede that it would be obvious to one of skill in the art to combine Komori et al. with Besling et al. and in fact note that the Office Action only asserts that Komori et al. and Besling et al. are analogous because they are in a similar field of endeavor and therefore it would be obvious to modify Komori et al. with Besling et al. with sparse further analysis. However, although Applicants traverse the obviousness analysis and believe that there are arguments which would lead one of skill in the art away from such combination, Applicants

believe that a clear case can be made that the current limitations are not taught by the combination of references.

Applicants first note that page 7 of the Office Action casts the first limitation of claim 1, 13 and 21 more broadly than the claim language itself. Notably, the Office Action states that “a memory that stores user voice data and data related to ...”. However, the claims actually cited a “user profile”. This is a subtle but important distinction and limitation in the claims that is unaddressed in the analysis. For example, the Office Action equates the parallel model combination taught in Komori et al. that utilizes noise models corresponding to Gaussian probability distribution as an “indication of a likelihood or probability of a user being in a particular environment.” The reason the limitation of a user profile should have weight in the analysis is that the claim requires the value associated with a probability of “the” user being in a particular acoustic environment. In other words, if a memory stores a user profile for Mary, Joe and Brian, each user profile will have a value associated with the probability of Mary, Joe or Brian being in a particular acoustic environment. These probability values are recited as being stored in a user profile and thus related to a particular user. The manner in which the Office Action casts these limitations generalizes this to be broad enough to cover the probability of any person being in any particular environment. This broadening is incorrect. The assertion that Besling et al. may reference a user ID does not remedy this deficiency of a user profile including the probability claimed.

Furthermore, Applicants respectfully traverse the fundamental assertion that the parallel model combination that uses Gaussian probability distributions relates or is “indicative” of a probability of a user being in a particular environment. Applicants submit that this is simply not the case. Applicants submit that when the details of the parallel model combination are studied it is clear that there is no reference to particular user profiles and any values associated with a

probability of a particular user being in a particular acoustic environment and stored in a user profile. For example, the Office Action references the Gales et al. article entitled "Robust Continuous Speech Recognition Using Parallel Model Combination." This article discusses in more detail the PMC scheme which has a goal of achieving noise robustness. In that scheme, experiments were used on the resource management (RM) database in which a 1,000 word continuous speech recognition task was studied. The database that they used for their experiments was the RM speaker independent task with either the Lynx Helicopter noise or Operation Room noise from the NOISEZ-92 database added. They explain in Section II that the PMC model uses a standard HMM with a Gaussian output probability distribution. The question is in the context of a standard HMM what does the Gaussian output probability distribution reflect? Does it provide any kind of indication of a likelihood of Mary, Joe or Brian being in a particular environment? The answer is clearly no. A hidden Markov model is basically a Markov chain where the output observation is a random variable X generated according to an output probabilistic function associated with each state.

In speech, the output probability distribution does not relate to the probability of the user being in a particular environment but it relates to the probability of a particular pattern in the observed sequence of a received speech utterance. In other words, when a user utters a certain sentence or a certain pattern of speech, what is the probability that those speech patterns match or correspond to training data associated with the HMM such that the speech can be appropriately identified. More information on the definition of in use of hidden Markov models can be found in "Spoken Language Processing", by Huang, Acero and Hahn, Prentice Hall PTR, 2001, Chapter 8. Applicants submit that it is a technically erroneous assertion by the Examiner to assert that the Gaussian probability distribution associated with the HMMs in the parallel model combination approach taught in Gales et al. and referenced in Komori et al. provides any

data related to the probability of a user being in a particular environment. The Gaussian output probability distributions only relate to distribution of speech either in a clean environment or in a noisy environment. For example, Section V of Gales et al. teach that there are several databases used in their study. The first was a clean speech model in the RM database. The second database was the NOISEX-92 database which includes a variety of noise sources, two of which were chosen for their experiments, the Lynx Helicopter noise, the power spectral density of which is shown in Fig. 2(a) and the operations room noise which is shown in Fig. 2(b).

Applicants respectfully note that in both of these background noise environments there is certainly nothing that is hinted at with regards to users being in the environment or a probability of a user being in a particular environment. In fact, Gales et al. and Komori et al. simply assume that 100% of the time a user is in a particular environment since that is where speech recognition occurs. This is because their approach to their study is quite simple with regards to a user being in a particular environment.

Gales et al. explain at the end of Section II that the basic PMC process involves receiving inputs to the scheme as (1) a clean speech model and (2) a noise model. This is shown in Fig. 1. Since the combination of speech and noise is most naturally expressed in the linear-spectral or log-spectral domains, it is simplest to model the effects of the additive noise on the speech parameters in one of these domains. They explain that the function that approximates this is referred to as the mismatch function. They then combine, according to the mismatch function, the clean speech models and the noise models and after the models have been combined, the estimate of the corrupted-speech model is transformed back into the cepstral domain if required. They explain the mismatch function in Section III.

Applicants note that their basic process therefore is to take a clean speech model and blend it according to a certain approach with a noise model (which includes noise associated

with a helicopter or noise associated with an operations room) and combine those and then perform speech recognition using a "corrupted speech HMM." As you can see, there is fundamentally no probability issue with regards to a user to whether a user is going to be in a particular environment. Gales et al. simply takes a chosen environment which may have helicopter noise or operations room noise and assumes that a user is going to be in that environment (100%) and produces the corrupted speech HMM for performing speech recognition in that environment. A simple example explains how the limitation of claim 1 differs from this approach. Again, assuming that Mary, Joe and Brian each have a user profile which includes a value associated with a probability user being in a particular environment, the profile for Brian may indicate from 7am to 9am he is on a train commuting to work and from 9:15am to 5pm he is in an office environment, and from 5:15 to 7:15 he is back on the train. In this regard, there may be a value associated with Brian being in a train or office environment at a particular time. This probability may be used as part of his user profile in compensating at least one speech recognition based on the user profile. For example, if the time is 2pm, a system may receive the user profile and compensate at least one speech recognition model with a high probability that the background associated with the speech recognition model may be an office environment. Such an approach with regards to probability is simply absent from the teachings of Komori et al. or Gales et al.

Applicants shall next discuss each portion of Komori et al. cited by the Examiner to further support our interpretation. First, the Examiner cites column 2, lines 25-34. This portion includes disclosing a basic speech recognition unit 203 for executing speech recognition processing for speech from a portable terminal 100. There is nothing in this portion other than basic information about the components of a speech recognition system. Column 3, lines 24-30 is also cited which teaches the environmental adaptation unit 201 that performs environmental

adaptation with respect to a speech recognition model in the speech recognition holding unit 205 on the basis of the environmental information to update the speech recognition model into the environment adaptation speech recognition model. This model is then held by the speech recognition model holding unit. Komori et al. explain in line 30 that the method for environmental adaptation may be the PMC technique used which is discussed above. Again, that approach basically involves taking a clean speech model and a noise model and combining them in the log-spectral domain in order to generate a corrupted speech model. There is nothing in this process that provides any data with regards to the limitation of a particular user having a user profile that includes a value associated with a probability of a user being in a particular acoustic environment. Next, the Office Action cites column 5, lines 3-35 which discusses adaptation data types. This portion merely discusses an embodiment of the invention which indicates of noise, "the average and variance of parameters in a noise interval are obtained and sent to the main body to perform noise adaptation of a speech recognition model by the PMC technique." The average and variance of parameters are simply the basic parameters associated with the distribution of noise in an environment and again have no bearing on a probability of a user being in a particular environment. In fact, in the Gales et al. discussion of the noise models they used helicopter noise and operations room noise which do not involve speech from people at all and thus have no bearing on any kind of probability of whether users would be in those environments, but merely are produced to reflect the characteristics (average and various of parameters) of the background noise.

Similarly, this portion of the reference also discusses a method of obtaining the average and variance of parameters in a speech interval of a certain duration and sending them to the main body and performing microphone characteristic adaptation of a speech recognition model by the CMS technique. This also again, is an approach of dealing with characteristics of a

microphone but Applicants respectfully submit that it provides no indication of a probability of Mary, Joe or Brian being in a particular environment. Thus, the concept of performing speaker adaptation of a speech recognition model is basically as discussed above relative to the function of the process of the PMC technique which is simply not "indicative" of a probability of a user being in a particular environment.

Next, the Office Action cites column 3, lines 6-16, in connection with the assertion that Gaussian probability distributions teach or suggest this limitation. Lines 6-16 of column 3 state the following:

"When the environment information is converted into a model in step S404, the flow advances to step S405 to cause the environment information creation unit 104 to create environment information. For the purpose of noise adaptation, for example, environment information is created by detecting a non-speech interval and obtaining the average and variance of parameters in the interval. For the purpose of microphone adaptation, environment information is created by obtaining the average and variance of parameters in a speech interval. For the purpose of speaker adaptation, a phonemic model or the like is created."

Again, this portion of the reference merely discusses processing the background environment (a non-speech interval) and then obtaining parameters associated with the microphone and blending that with data for the purpose of speech adaptation. This analysis is silent regarding the probability of the user being in a particular environment and simply assumes that there will be a particular environment and a speaker will be speaking in that environment. The rest of column 3 also supports our position. Next, the Office Action recites column 4, lines 1-6 which states that the number of quantization points may be decreased by a method of merging all distributions into one distribution, searching 3σ (e.g., a range in which most of samples appearing in a Gauss distribution are included) for the maximum and minimum values, and dividing the interval therebetween into equal portions. The number of quantization points here discussed and the Gaussian distribution discussed again are in the context of speech recognition and relates to the distribution of speech in an environment and assumes that a user

will speak in the environment. There is no data that provides any indication of a probability of a user being in a particular environment.

Next, column 5, lines 1-47 is cited (column 5, lines 3-35 were cited above). As previously noted, this portion of the reference merely discusses the basic use of average and variance parameters in the context of analyzing probabilities of noise, microphone characteristics and speaker characteristics. Lines 36-47 teach the following:

“In this embodiment, the initialization mode is to be performed before the speech recognition mode. Once the initialization mode is completed, however, speech recognition can be resumed from the speech recognition mode under the same environment. In this case, the previous environment information is stored on the portable terminal 100 side, and environment information created in resuming speech recognition is compared with the stored information. If no change is detected, the corresponding notification is sent to the main body 200, or the main body 200 performs such determination on the basis of the sent environment information, thus executing speech recognition.”

Applicants note that the Office Action on page 5 references this particular portion of Komori et al. by stating that it teaches a comparison of current noise environment to previous noise environment models as apparently fundamental to the assertion that the probability distribution in Komori et al. are indicative of a probability of a user being in these particular environments. However, when the actual steps that are outlined on lines 36-47 are traced, it is clear that they provide no data or information that might be indicative of a probability of a user being in a particular environment. Before speech recognition occurred, an initialization mode is completed. Once the initialization mode is complete speech recognition may be resumed in a speech recognition mode. While there may be a comparison of previous environmental information and environmental information that is created in resuming speech recognition, in each case, speech recognition occurs assuming that the user is in the environment. There is simply no mention of a probability of the user being in the environment and it is simply assumed that the user is speaking in the environment. The purpose of this portion of Komori et al. simply

appears to be an opportunity of using the environmental information both for speech recognition processing and an improvement in speech efficiency. See lines 48-50. As can be seen, it is too much of a stretch to read into this use of background information for speech recognition that there is any kind of analysis involving a probability of whether the user is going to be in a particular environment. One reason why the probability issue has no bearing in Komori et al. is the process outlined in paragraph 5 involves a user already in a particular environment and essentially dynamically retrieving information about that background environment prior to speech recognition. This is the initialization mode that is performed prior to speech recognition. Once they complete initialization, then they resume speech recognition in that environment. The scenario here is simple to determine. If a user has a mobile device that performs speech recognition and is in a first environment, the initialization mode will identify the background environment noise prior to speech recognition and store that information. If a user moves to a different environment, then prior to speech recognition the initialization mode is performed and the environment is tested. If no change is detected, i.e., the background environment is the same, then a notification is sent to the main body and the system performs speech recognition on the sent environmental information without any change. This approach again differs from the concept of having a user profile which includes a value associated with a probability of a user being in a particular acoustic environment because this approach merely takes each environment as it comes and performs the initialization mode prior to speech recognition on the current environment. If there is a change from a previous environment then adjustments are made, if there is no change detected then no change in the background noise environment is made. Thus, there is no disclosure or hint of the probability value of the present claims.

Finally, column 4, lines 1-25 of Komori et al. are cited for a plurality of stored noise models. This approach taught in column 4, lines 1-25, are the steps that occur prior to the speech

recognition mode in step S412 of Figure 2. This portion of the reference in column 4 involves a detailed description of steps S409, S410 and S411 which involve creating a scaler quantization table for speech recognition communication and transmitting the created scaler quantization table to a portable terminal 100. The distribution discussed in this portion of the reference (which really begins at column 3, line 55) involves a distribution of environment adaptive speech recognition models. Thus, according to the method, the scaler quantization table of the respective dimensions is created by using "the distribution of environmental adaptive speech recognition models." The purpose is that the bit rate for communication can be decreased without degrading a recognition performance and allowing efficient communication. See column 4, lines 11-14. It appears that the actual Gaussian distribution discussed in the this portion of the reference is actually a Gaussian distribution of environmental adaptive speech recognition models, rather than anything that is indicative of a probability of a user being in a particular environment. At a higher level, Applicants would certainly traverse the Examiner's analysis by noting that the particular reference to the Gaussian distribution is in the context of steps S409, S410 and S411 shown in Figure 2 which involve the creation of the scaler quantization tables in a portable terminal which have as their purpose a goal of improving the communication efficiency of speech recognition. This is simply far a field from the probability of a user being in a particular environment with a value in a user profile for that probability.

Accordingly, Applicants have explained that the Gaussian distribution discussed in the references has no bearing on the probability of a user being in a particular environment.

Applicants also note in response to the summary on page 6 of the Office Action with regard to issue (c) that claims 1, 13 and 21 have been amended to remove the alternative approach and thus require the limitation of the probability value being associated with a user being in a particular environment. Point issue (b) is addressed extensively above with regards to

whether or not Komori et al. teaches the claimed feature of the probability value being associated with “the user being in a particular acoustic environment”. With regards to point (a), Applicants refer to the beginning of our Remarks wherein we highlight that this requirement in claim 13 and claim 21 focuses on compensating the speech recognition model for recognizing the speech utterance based on the user profile data which in point (a) is asserted to have “changed the claim scope” but which ironically was language not used on page 7 of the Office Action. The user profile must include information associated with the probability of *that* user being in a particular environment which is a value not found in the Besling et al. adaptation profile that may be identified by a user ID in column 7, line 35 – column 8, line 55. Accordingly, Applicants respectfully submit that claims 1, 13 and 21 are patentable and in condition for allowance. Claims 2-4 and 8-9 each depend from claim 1 and recite further limitations therefrom. Accordingly, Applicants respectfully submit that these dependent claims are patentable as well.

Claims 15 and 18-19 each depend from claim 13 and recite further limitations therefrom and therefore are patentable and in condition for allowance. Claims 22-23 and 27-29 each depend from claim 21 and recite further limitations therefrom and accordingly are patentable and in condition for allowance.

Rejection of Claims 5, 6 and 25 Under 35 U.S.C. §103(a)

The Office Action rejects claims 5, 6 and 25 under 35 U.S.C. §103(a) as being unpatentable over Komori et al. in view of Besling et al. and further in view of Kanevsky et al. (U.S. Patent No. 6,442,519) (“Kanevsky et al.”). Applicants respectfully traverse this rejection and do not acquiesce that one of skill in the art would have sufficient motivation or suggestion to combine Komori et al., Besling et al. and Kanevsky et al. However, inasmuch as claims 5 and 6 depend from claim 1 and claim 25 depends from claim 21, Applicants submit that these claims are patentable and in condition for allowance.

Rejection of Claims 7 and 26 Under 35 U.S.C. §103(a)

The Office Action rejects claims 7 and 26 under 35 U.S.C. §103(a) as being unpatentable over Komori et al. in view of Besling et al. and further in view of Hunt et al. (U.S. Patent No. 6,094,476) ("Hunt et al."). Applicants respectfully traverse this rejection and do not acquiesce that one of skill in the art would combine Komori et al., Besling et al. and Hunt et al. However, inasmuch as claim 7 depends from claim 1 and claim 26 depends from claim 21, Applicants submit that these dependent claims are patentable and in condition for allowance.

Rejection of Claims 10, 17 and 30-31 Under 35 U.S.C. §103(a)

The Office Action rejects claims 10, 17 and 30-31 under 35 U.S.C. §103(a) as being unpatentable over Komori et al. in view of Besling et al. and further in view of Heck et al. (U.S. Patent No. 5,950,157) ("Heck et al."). Applicants respectfully traverse this rejection and do not acquiesce that one of skill in the art would have sufficient motivation or suggestion to combine Komori et al., Besling et al. and Heck et al. However, inasmuch as claim 10 depends from claim 1, claim 17 depends from claim 13 and claims 30-31 depend from claim 12, Applicants submit that these claims are patentable and in condition for allowance.

Rejection of Claims 11-12, 20 and 32 Under 35 U.S.C. §103(a)

The Office Action rejects claims 11-12, 20 and 32 under 35 U.S.C. §103(a) as being unpatentable over Komori et al. in view of Besling et al. and further in view of Cilurzo et al. (U.S. Patent No. 6,434,526) ("Cilurzo et al."). Applicants respectfully traverse this rejection and do not acquiesce that one of skill in the art would have sufficient motivation or suggestion to combine Komori et al. with Besling et al. and Cilurzo et al. However, inasmuch as claims 11 and 12 each depend from claim 1, claim 20 depends from claim 13 and claim 32 depends from claim 21, Applicants submit that these claims are patentable and in condition for allowance.

Rejection of Claim 14 Under 35 U.S.C. §103(a)

The Office Action rejects claim 14 under 35 U.S.C. §103(a) as being unpatentable over Komori et al. in view of Besling et al. and further in view of Ranzino (U.S. Patent No. 6,281,811) ("Ranzino"). Applicants traverse this rejection and do not acquiesce that one of skill in the art would have sufficient motivation or suggestion to combine Komori et al., Besling et al. and Ranzino.

Applicants respectfully submit that one of skill in the art would not have sufficient motivation or suggestion to combine these references because Razino only teaches a very generic keyword recognition approach. Razino's invention is a communication and location system that provides a user with the means of locating him or herself in an area such as a shopping mall using voice commands. The voice commands are taught in column 4 in which the user speaks into a headset and says words such as "food", "restaurants", "snacks" and the system will recognize these keywords and determine that the user is interested in a food resource. This reference is clearly disclosed at a very high level with regards to voice recognition and an objective reading indicates that it is only used for the purpose of simple keyword recognition.

Applicants would submit that it is non-analogous to Besling et al. which disclose a pattern recognition system for use in large vocabulary continuous speech recognition systems or handwriting recognition systems which typically use a collection of recognition models to recognize an input pattern. See column 1, lines 50-53. Their figure 1 illustrates a structure of a large vocabulary continuous speech recognition and throughout the reference they disclose a plurality of different speech recognition models for the purpose of large vocabulary continuous speech recognition. See column, 3, lines 45-46; column 4, lines 24-25, lines 58-60; column 5, lines 35-45; column 7, lines 45-49, etc. Accordingly, Applicants respectfully submit that one of skill in the art would not likely view Razino as a reference that would be blended with the

distributed pattern recognition system of Besling et al. Therefore, Applicants submit that one of skill in the art would not combine these references for common sense reasons. Furthermore, claim 14 recites the step of a controller that identifies a mobile device by a radio frequency identification tag. Applicants traverse that Razino teaches this particular limitation inasmuch as column 4, lines 31-42 only teach that the user makes a verbal request by speaking into a headset microphone and the request is transmitted with the "user's identification code" to identify the user by means of the user's identification code. There is no reference to identifying a mobile device by using a radio frequency identification tag. This is also confirmed by the Examiner's own analysis on page 15 of the Office Action in which the Office Action states that Razino is blended with the other references "to provide a means for easily identifying a particular user to communicate information related to that user's preferences." This is simply not the limitation of claim 14 and thus, Applicants submit that claim 14 is patentable and in condition for allowance.

Rejection of Claim 16 Under 35 U.S.C. §103(a)

The Office Action rejects claim 16 under 35 U.S.C. §103(a) as being unpatentable over Komori et al. in view of Besling et al. and further in view of Byers (U.S. Patent No. 6,219,645) ("Byers"). Applicants traverse this rejection and do not concede that one of skill in the art would have sufficient motivation or suggestion to combine Byers with Komori et al. and Besling et al. However, inasmuch as claim 16 depends from claim 13 and recite further limitations therefrom, Applicants submit that this claim is patentable and in condition for allowance.

Rejection of Claim 24 Under 35 U.S.C. §103(a)

The Office Action rejects claim 24 under 35 U.S.C. §103(a) as being unpatentable over Komori et al. in view of Besling et al. and further in view of Sonmez et al. (U.S. Patent No. 5,745,872) ("Sonmez et al."). Applicants traverse this rejection and do not acquiesce that it would be obvious for one of skill in the art to combine Komori et al., Besling et al. with Sonmez

et al. However, inasmuch as claim 24 depends from claim 21, and recite further limitations therefrom, Applicants submit that this claim is patentable and in condition for allowance.

CONCLUSION

Having addressed all rejections and objections, Applicants respectfully submit that the subject application is in condition for allowance and a Notice to that effect is earnestly solicited. If necessary, the Commissioner for Patents is authorized to charge or credit the **Novak, Druce & Quigg, LLP, Account No. 14-1437** for any deficiency or overpayment.

Respectfully submitted,

Date: November 14, 2007

By: 

Correspondence Address:

Thomas A. Restaino
Reg. No. 33,444
AT&T Corp.
Room 2A-207
One AT&T Way
Bedminster, NJ 07921

Thomas M. Isaacson

Attorney for Applicants
Reg. No. 44,166
Phone: 410-286-9405
Fax No.: 410-510-1433